

## The Convolution Recurrent Network of Using Residual LSTM to Process the Output of the Downsampling for Monaural Speech Enhancement

**Authors :** Shibo Wei, Ting Jiang

**Abstract :** Convolutional-recurrent neural networks (CRN) have achieved much success recently in the speech enhancement field. The common processing method is to use the convolution layer to compress the feature space by multiple upsampling and then model the compressed features with the LSTM layer. At last, the enhanced speech is obtained by deconvolution operation to integrate the global information of the speech sequence. However, the feature space compression process may cause the loss of information, so we propose to model the upsampling result of each step with the residual LSTM layer, then join it with the output of the deconvolution layer and input them to the next deconvolution layer, by this way, we want to integrate the global information of speech sequence better. The experimental results show the network model (RES-CRN) we introduce can achieve better performance than LSTM without residual and overlaying LSTM simply in the original CRN in terms of scale-invariant signal-to-distortion ratio (SI-SNR), speech quality (PESQ), and intelligibility (STOI).

**Keywords :** convolutional-recurrent neural networks, speech enhancement, residual LSTM, SI-SNR

**Conference Title :** ICSLP 2021 : International Conference on Speech and Language Processing

**Conference Location :** San Francisco, United States

**Conference Dates :** November 01-02, 2021